TTL-IQA: Transitive Transfer Learning based No-reference Image Quality Assessment

Xiaohan Yang, *Student Member, IEEE,* Fan Li, *Member, IEEE,* and Hantao Liu, *Member, IEEE*

***Abstract*—Image quality assessment (IQA) based on deep learning faces the overfitting problem due to limited training samples available in existing IQA databases. Transfer learning is a plausible solution to the problem, in which the shared features derived from the large-scale Imagenet source domain could be transferred from the original recognition task to the intended IQA task. However, the Imagenet source domain and the IQA target domain as well as their corresponding tasks are not directly related. In this paper, we propose a new transitive transfer learning method for no-reference image quality assessment (TTL-IQA). First, the architecture of the multi-domain transitive transfer learning for IQA is developed to transfer the Imagenet source domain to the auxiliary domain, and then to the IQA target domain. Second, the auxiliary domain and the auxiliary task are constructed by a new generative adversarial network based on distortion translation (DT-GAN). Furthermore, a TTL network of the semantic features transfer (SFTnet) is proposed to optimize the shared features for the TTL-IQA. Experiments are conducted to evaluate the performance of the proposed method on various IQA databases, including the LIVE, TID2013, CSIQ, LIVE multiply distorted and LIVE challenge. The results show that the proposed method significantly outperforms the state-of-the-art methods. In addition, our proposed method demonstrates a strong generalization ability.**

***Index Terms*—Transitive transfer learning, image quality assessment, auxiliary domain, distortion translation, semantic feature transfer, generative adversarial network.**

I. Introduction

With the fast development of social media and the increasing demand for imaging services, a large number of digital images are generated, stored, processed and transmitted every day [1]. Through these different stages of the imaging pipeline, image signals are subject to a wide variety of distortions, which may result in visual quality degradation. A reliable IQA method can help quantify the image quality on the Internet and accurately assess the performance of image processing algorithms from the perspective of human observers. Therefore, it is crucial to develop effective IQA methods.

Objective IQA methods are classified in general into three categories depending on the availability of the reference image: full-reference IQA (FR-IQA) [2], [3], reduced-reference IQA (RR-IQA) [4], [5], and no-reference IQA (NR-IQA). However, since the reference is not accessible in many practical scenarios, NR-IQA attracts a significant amount of research interests in recent years.

Most of the traditional NR-IQA methods commonly adopt some handcrafted features of the distorted images, and then train a shallow regression model (e.g., support vector regression) to map the feature representations to subjective quality scores [6]-[8]. An obvious limitation of those NR-IQA methods is that the handcrafted features may not be powerful enough to adequately represent complex structures and distortions of images for the IQA task. With the great success of deep learning in the field of image recognition and processing, the deep learning method provides a very promising strategy for addressing the challenging NR-IQA problem [12]. This is because the remarkable capability of the deep neural network (DNN) in automatically discriminating features related to image quality. Nevertheless, the success of deep learning methods relies heavily on large-scale annotated data, such as the Imagenet dataset for the image recognition task [14]. Unfortunately, for the IQA task, there does not exist a large database of training images with the groundtruth labels of human subjective quality scores.

Therefore, researchers pay more attention to the use of a variety of data enhancement methods to generate more training samples for IQA task [15]. The most popular method is to divide an image into small image patches. The subjective score of the whole image or the proxy score derived from an FR metric is used as the groundtruth label of each image patch. However, these groundtruth labels are inaccurate to represent the real subjective scores of image patches. Some methods aim to transfer the shared features from the large-scale Imagenet source domain to the IQA target domain to complete the IQA quality score task, which can reduce the burden of training a DNN model with an IQA database from scratch [16], [17]. However, since the Imagenet source domain and the IQA target domain and their respective tasks are not directly related, finding ways to achieve an effective transfer learning approach remains challenging.

In this paper, we propose a transitive transfer learning based no-reference image quality assessment method (TTL-IQA), which aims to identify and reduce the irrelevance between the Imagenet source and the IQA target domains and tasks. Our contributions are summarized as follows.

(1) We develop an architecture of the multi-domain transitive transfer learning (TTL) for IQA. An auxiliary domain is designed to act as the intermediate bridge between the Imagenet source and the IQA target domains, which aims

Xiaohan Yang and Fan Li are with the Ministry of Education Key Laboratory for Intelligent Networks and Network Security, School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an, 710049, China. (e-mail: yangxiaohan@stu.xjtu.edu.cn; lifan@mail.xjtu.edu.cn).

Hantao Liu is with the School of Computer Science and Informatics, Cardiff University, Cardiff, CF243AA, U.K. (e-mail: LiuH35@cardiff.ac.uk).

This research work was supported in part by National Science Foundation of China (62071369). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Marco Carli. *(Corresponding author: Fan Li.)*

to enhance the multi-domain correlation by associating multi-domain image properties. Also, the auxiliary task is designed to act as the intermediate bridge between the recognition and the IQA tasks, which aims to enhance the multi-task correlation by associating multi-task labels.

(2) We construct the auxiliary domain and the auxiliary task by a new generative adversarial network based on distortion translation (DT-GAN). For the auxiliary domain construction, the hallucinated distortion images are generated by the DT-GAN using source images that are not from the IQA databases. Meanwhile, a stack connected resize convolution module is designed to perform the distortion distribution translation from the IQA images to the hallucinated distortion images. For the auxiliary task, the quality-level labeling strategy is proposed to generate the labels for the hallucinated distortion images.

(3) We propose a new TTL network of the semantic features transfer (SFTnet) to optimize the shared features for the TTL-IQA. The semantic discrimination adaptation (SDA) block is designed as the novel attention unit to adaptively enhance the useful shared features of the IQA task relevant to the multi-domain and multi-task learning, while suppressing the useless features, including discriminative and class-specific features.

II. RELATED WORK

In this section, we provide a review of the recent NR-IQA methods. A more detailed review of the NR-IQA methods can be found in [12], [15], [27].

A. The traditional NR-IQA methods

The traditional NR-IQA methods attempt to extract some specific features that could discriminate distortion images from the pristine images, and learn a shallow regression model to map the image representations onto scalar quality scores. The first category methods [6]-[8] extract natural scene statistics (NSS) as features based on the statistical regularity of natural images. However, it remains limited for these handcrafted features to fully represent complex image structures and distortions. The second category methods [9]-[11] extract features by feature encoding with respect to a learned codebook. The MSDD method [10] first extract compact and discriminative quality-aware features of local image patches by using FR method to optimize discriminative dictionary. Then, the image level features are aggregated and the SVR method is used to predict image quality. However, the FR method is difficult to extract discriminative quality-aware features of authentic distortion images and the SVR method is a shallow regression model, which is difficult to simulate the complex perception mechanism of humans [15]

B. Deep learning methods for NR-IQA

In recent years, much works has used deep learning for NR-IQA. The motivation is that the DNN can automatically capture more deep features relevant to quality assessment and so to improve prediction performance. However, challenges remain for the deep learning methods for NR-IQA, primarily due to the lack of sufficient IQA databases.

To address this problem, there are two methods to enhance the labeled image data for the IQA task. One is the image patch-based method [28]-[30]. In this approach, each image of the IQA database is divided into a large number of image patches to achieve the internal enhancement. Kang et al. [28] first divided an image into several image patches and used the subjective score of the whole image as the label for all patches to train the DNN model. Then, the image quality is predicted by using the average score of all image patches. Instead of using the subjective score as the label for all image patches, some methods use the FR metric as the proxy patch label [29], [30].

The other method is based on transfer learning[18], [31]-[33]. In this approach, the pre-trained VGG network (VGG) for the recognition task is commonly transferred the shared features derived from the large-scale non-IQA image databases to achieve the IQA task. In [18], the VGG is used to transfer the shared features from the recognition task to the IQA task. Moreover, some images in the image recognition datasets are simulated artificially using typical distortion types in the IQA databases. In [31], the method expands some external images through some functions relevant to specific distortion types in the IQA database and the corresponding quality labels can be obtained by varying the parameters in the distortion functions. Then, the Siamese network contained with twin pre-trained VGG [34] derived from the recognition task is used to rank these external images quality level, and then fine-tune a branch of the VGG to assess image quality score. Similarly, Zhang et al. [32] use the same method to expand IQA database and train the DNN and the pre-trained VGG derived from the recognition task to predict image quality both the synthetic and authentic distortion images. In addition, some methods opt to the GAN [35], [36] to generate the hallucinated reference images constrained on the distortion images in the IQA database and use the DNN model to predict image quality [33].

Compared with different NR-IQA methods, we summary the difference between our method and different NR-IQA methods. Different from the tradition MSDD method [10], Our TTL-IQA method is suitable for evaluating quality of images with authentic, mixed and synthetic IQA databases and the quality prediction accuracy can be enhanced by using deep learning. In contrast to the method in [18], our TTL-IQA method enhances the relationship between the ImageNet source domain and the IQA target domain as well as their corresponding visual tasks by constructing the auxiliary domain and auxiliary task. For the methods in [31], [32], they can only use simple functions to simulate some synthetic/artificial distortions in images with specific and known distortion types, and they cannot simulate e.g., authentic and mixed distortions in images. Also, the method in [33] relies on the information from the pristine reference images, therefore, is not suitable for the evaluation of authentic distortions in images. Our TTL-IQA method overcomes these problems by using GT-GAN to simulate a variety of distorted images, including synthetic, authentic, and mixed distortions in images. In addition, our TTL-IQA method does not highly depend on the pristine reference images in the IQA database.

III. The Framework of TTL-IQA

A. The limitation of the transfer learning method for IQA

The transfer learning method is to use the shared features of the VGG from the large-scale Imagenet source domain for the IQA target domain to achieve the task transfer from the recognition task to the IQA quality score task. The following abbreviations are used in this paper, as shown in Table I.

The definitions of the Imagenet source domain and the IQA target domain are given as:

$$D_s = \{\chi_s, P(X_s)\} \tag{1}$$

$$D_t = \{\chi_t, P(X_t)\} \tag{2}$$

where D_s and D_t are the Imagenet source domain and the IQA target domain, respectively. χ_s and $P(X_s)$ are the image feature space and marginal probability distribution in D_s, respectively. X_s is a image set in D_s, which belongs to χ_s. χ_t and $P(X_t)$ are the image feature space and marginal probability distribution in D_t, respectively. X_t is a image set in D_t, which belongs to χ_t.

The recogniton task and the IQA quality score task are defined as:

$$T_s = \{y_s, f_s(\cdot)\} \tag{3}$$

$$T_t = \{y_t, f_t(\cdot)\} \tag{4}$$

where T_s and T_t are the recogniton task and the IQA quality score task, respectively. y_s and y_t are the classification and the quality score labels, respectively. $f_s(\cdot)$ and $f_t(\cdot)$ denote the predictive functions for the classification and the IQA quality score tasks, respectively.

The major limitations of the transfer learning method for IQA are described below, and as shown in Fig. 1.

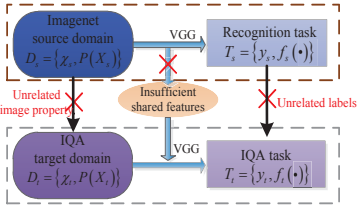

Fig. 1. The limitations of the transfer learning method for IQA.

(1) D_s and D_t are not directly related. Although χ_s and χ_t are similar, $P(X_s)$ and $P(X_t)$ are not similar, because there is a little overlap in the image properties between the two domains. The image properties in D_s represent salient object, shapes, size, activities and so on [20], [21]. Nevertheless, the image properties mainly include distortion, image content and salient object [22], [23] in D_t.

(2) T_s and T_t are dramatically unrelated. First, y_s and y_t are irrelevant. y_s is the object classification and y_t is the quality scoring. Moreover, $f_s(\cdot)$ and $f_t(\cdot)$ are different. For T_s, $f_s(\cdot)$ aims to learn the mapping relationship between the images of the Imagenet dataset and object classification labels. However, for T_t, $f_t(\cdot)$ is to map the relationship between distortion images and quality scores.

B. The difficulities of TTL method

In order to overcome the irrelevance between D_s and D_t, as well as T_s and T_t, it is necessary to construct an auxiliary domain and its task to enhance the correlation between D_s and D_t, as well as T_s and T_t.

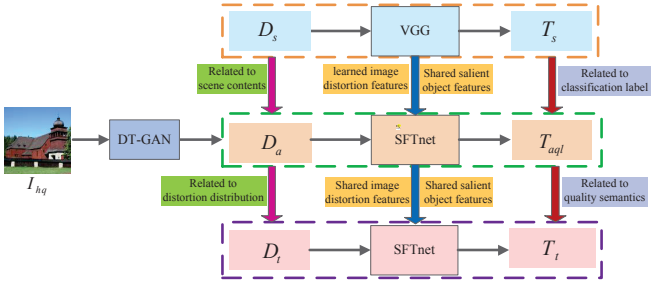

Fig. 2. The framework of the proposed TTL-IQA for IQA.

The requirements for constructing D_a and T_a are summarized as follows.

(1) The construction of D_a needs to associate the marginal probability distribution of image data between D_s and D_t. It means that the marginal probability distribution of image data is mainly related to the common image properties and the important image properties for the IQA task. Since D_s and D_t have the constraints of rich scene content, it can act as the characteristic in D_a to associate the common image properties. Furthermore, since the distortion is the most important property to the IQA image data, the images in D_a need to contain various distortion features.

(2) The construction of T_a needs to associate the labels. For multi-task labels, the label of T_a not only needs to associate the classification label for T_s, but also to maximize the transition to perceived quality label for T_t.

(3) The appropriate TTL network needs to be designed to associate the shared features of prediction functions between the two tasks. For T_s, the salient object and the highly discriminative and class-specific features are used to construct the prediction function. However, for T_a and T_s, the distortion and salient object are most useful features to construct the prediction functions. Therefore, the appropriate TTL network needs to share these useful features to construct the multi-task correlation.

However, there are some difficulties to construct appropriate D_a and T_a to complete TTL. First, the construction of D_a is incomplete by using the popular artificial distortion simulation methods [31], [32]. In fact, there is a lack of the pristine reference images and the prior distortion information especially for the authentic and the mixed IQA databases [37], [51]. Second, the labels between multiple tasks (T_s,T_a,T_t) are not highly related, because the labels in T_a cannot be highly associated with multiple tasks. Third, the shared features are insufficient from T_s to T_a. Although the VGG is useful for transfer learning, it not only transfers the shared features related to salient object that are useful to T_s and T_{aql}, but also transfers the shared highly discriminative and class-specific features that are useful for T_s but useless for T_{aql}. It makes the performance of transfer learning is non-optimal.

TABLE I
THE ABBREVIATIONS USED IN THIS PAPER

abbreviations	Symbols	abbreviations	Symbols
Imagenet source domain	D_s	Recognition task	T_s
Auxiliary domain	D_a	Auxiliary task	T_a
–	–	Auxiliary quality level task	T_{aql}
IQA target domain	D_t	IQA quality score task	T_t
The IQA distortion image	I_d	The hallucinated distortion image	$I_d^{'}$
The high quality image	I_{hq}	The hallucinated high quality image	$I_{hq}^{'}$
The generative network of the hallucinated distortion quality images	D-Gnet	–	–
The discriminative network of the hallucinated distortion quality images	D-Dnet	–	–
The generative network of the hallucinated high-quality images	H-Gnet	–	–
The discriminative network of the hallucinated high-quality images	H-Dnet	–	–

C. The framework of the proposed TTL-IQA for IQA

Therefore, we propose a new TTL-IQA framework, as shown in Fig. 2. The DT-GAN is proposed to generate large-scale hallucinated distortion images and corresponding quality semantic labels. Meanwhile, the SFTnet is proposed to optimize the shared features to improve prediction performance of image quality for TTL.

The advantages of the proposed DT-GAN and SFTnet are as follows, respectively.

(1) The proposed DT-GAN aims to construct D_a, which contains large-scale hallucinated distortion images. Compared with the traditional simulation distortion approaches [31], [32], DT-GAN can easily learn to simulate various distortion types, including synthetic, mixed and authentic distortions using DT-GAN architecture and its loss functions. This method is widely applicable and can overcome the limitations of other simulation methods [31], [32] that require explicit noise functions. Also, DT-GAN can learn the distribution of distortion properties of the IQA databases, making sure the simulated distortions have a similar distribution to that of the image distortions in the IQA databases.

Therefore, once D_a is constructed, domain correlation is enhanced. This is because from D_s to D_a, the properties of various image scenes in D_a are associated with the common image properties of D_s and D_t. From D_a to D_t, the distortion properties in D_a can act as the specific IQA property to associate D_t, which can overcome the disadvantage of artificial distortion simulation methods.

(2) The proposed DT-GAN aims to construct T_{aql}, including quality semantic labels . This is because DT-GAN can simultaneously generate quality labels of hallucinated distortion images. Once T_{aql} is constructed, task correlation is enhanced. From T_s to T_{aql}, the labels are about the classification tasks. From T_{aql} to T_t, the labels are directly related to the quality semantics, which aim to describe the image quality from the coarse-grained quality level to the fine-grained quality score.

(3) The proposed SFTnet aims to enhance the shared features in the TTL process. When transferring from T_s to T_{aql}, the SFTnet can enhance the shared salient object features, which is useful for both T_s and T_{aql}, and also it suppresses the shared highly discriminative and class-specific features, which is useful for T_s but useless for T_{aql}. When transferring from T_{aql} to T_t, it is easy to inherit all the shared features obtained from T_{aql} to achieve T_t. Since the shared salient object features obtained from T_s and the learned distortion features obtained from T_{aql} are most useful to achieve T_t, the SFTnet is only to be fine-tuned to achieve the transformation from T_{aql} to T_t.

IV. The Construction of the Auxiliary Domain and the Auxiliary Task for TTL-IQA

In this part, we introduce the architecture of the proposed DT-GAN in detail, which aim to construct the auxiliary domain and the auxiliary quality level task for TTL-IQA.

A. The architecture of DT-GAN

In order to construct the auxiliary domain D_a and the auxiliary quality level task T_{aql}, we propose a DT-GAN architecture to generate a large number of hallucinated distortion images with three quality levels, whose distortion distribution is similar to that of the IQA databases. In Fig. 3, we give the architecture of the proposed DT-GAN.

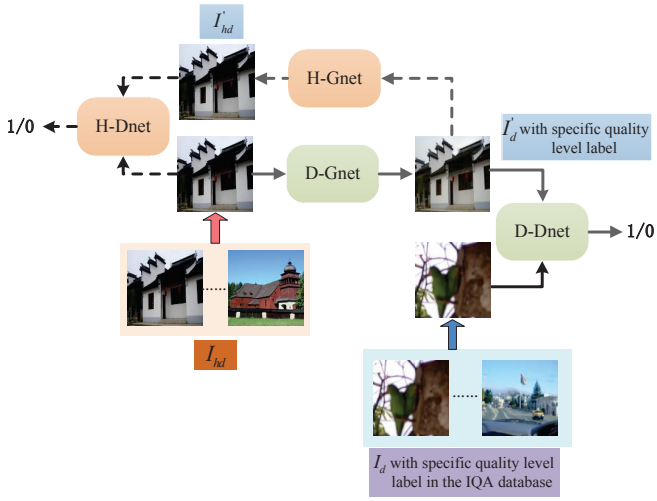

Fig. 3. The architecture of the proposed DT-GAN.

1) The composition of the DT-GAN: The DT-GAN consists of four parts, including the generative network of the hallucinated distortion images (D-Gnet), the discriminative network of the hallucinated distortion images (D-Dnet), the generative network of the hallucinated high-quality images (H-Gnet), and the discriminative network of the hallucinated high-quality images (H-Dnet).

The D-Gnet aims to generate the hallucinated distortion images $I_d^{'}$ from the large number of the high quality images I_{hq} beyond the contents of the IQA databases. Furthermore, the D-Dnet is trained by using $I_d^{'}$ and the distorted images I_d with a specific quality level in the IQA databases. The goal is to translate the distortion distribution from I_d to $I_d^{'}$ so that the distortion distribution of $I_d^{'}$ is indistinguishable with that of I_d.

TABLE II
THE PARAMETER SETTINGS IN DT-GAN

DT-GAN	L	KS	S	P	IC	OC	U
D-Gnet/H-Gnet	SC1	7	1	0	3	64	×
	SC2	3	2	1	64	128	×
	SC3	3	2	1	128	256	×
	RB1	3	1	0	256	256	×
	RB2	3	1	0	256	256	×
	RC1	3	1	0	256	128	NN
	RC2	3	1	0	256	128	NN
	SC4	7	1	0	64	3	×
D-Dnet/H-Dnet	SC_{m1}	4	2	1	3	64	×
	SC_{m2}	4	2	4	64	128	×
	SC_{m3}	4	2	1	128	256	×
	SC_{m4}	4	1	1	256	512	×
	SC_{m5}	4	1	1	512	1	×

However, the disadvantage is the distortion translation is highly under-constrained by using the D-Dnet and the D-Gnet. Since the image scenes of I_{hq} are different from that of I_d, this leads to the contents of $I_d^{'}$ to be destroyed to meet the image distribution consistency between $I_d^{'}$ and I_d.

In order to ensure that the content distribution remain unchanged for $I_d^{'}$, the H-Gnet aims to construct an inverse mapping relationship that can translate $I_d^{'}$ into the hallucinated high-quality images $I_{hq}^{'}$, which is similar to the content of I_{hq}. It prevents the interference of the image content of the IQA database to $I_d^{'}$. Finally, $I_{hq}^{'}$ is to fool the H-Dnet, which cannot distinguish the content difference between $I_{hq}^{'}$ and I_{hq}.

2) The architectures of D-Gnet and H-Gnet in GT-GAN: In the DT-GAN, the D-Gnet and the H-Gnet architectures are the same, as shown in Fig. 4. It follows the auto-encoder and decoder modules. The architecture of auto-encoder module consists of three standard convolution (SC) layers and three stacked residual blocks (RB) [42], which aims to obtain the deep features of the lower dimensions and avoid the gradient vanish of the deep network. The architecture of decode module consists of two resize convolution (RC) layers and a SC layer. It aims to alleviate the external checkerboard distortion in the decoder module caused by the general deconvolution (DC) operation [43],[44].

The parameter settings of the D-Gnet or the H-Gnet are shown in Table II. Note that L means the different layers and the KS means the kernel size. The S, P, IC, OC and U mean the stride, padding, input channel, output channel and upsampling, respectively.

3) The architectures of D-Dnet and H-Dnet in GT-GAN: In the DT-GAN, the D-Dnet and the H-Dnet architectures are also the same, as shown in Fig. 5. It contains the five SC layers and the parameter settings are shown in Table II. The D-Dnet aims to discriminate the fake $I_d^{'}$ from the real I_d. Meanwhile,

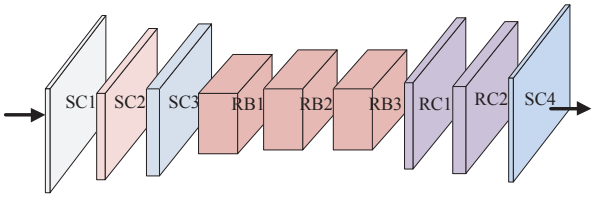

Fig. 4. The D-Gnet/H-Gnet architecture.

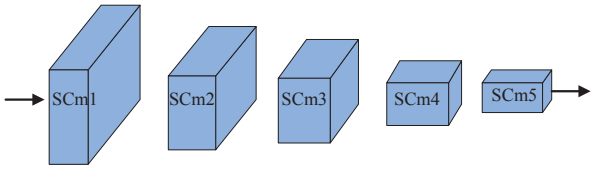

Fig. 5. The D-Dnet/H-Dnet architecture.

the H-Dnet aims to discriminate the fake $I_{hq}^{'}$ from the real I_{hq}.

B. The quality level label of the hallucinated distortion image for training DT-GAN

After the $I_d^{'}$ image are generated to construct D_a, the reasonable labels of $I_d^{'}$ need to be designed to construct T_a for TTL. It not only needs to associate the labels from T_s to T_t, but also associate the labels from T_a to T_t. The most reliable method is to evaluate the quality label of $I_d^{'}$ from the large number of subjects. However, this method is time-consuming and impractical. Therefore, we propose the quality level strategy to roughly generate the label of $I_d^{'}$ for training DT-GAN.

The definition of quality level is to describe quality degradation. It aims to classify the subjective scores of the distortion images into three quality levels, including High (i.e., with perceptible but not annoying artifacts), Medium (i.e., with noticeable and annoying artifacts) and Low (i.e., with very annoying artifacts) quality levels for each IQA database. Especially, each quality level includes the distortion images with similar subjective scores and the number of images in the three quality levels is approximately the same, as shown in Table III.

Let $\vec{S}$ be the subjective scores of the distortion images in each IQA database. $\vec{S_1}$=$sort\{\vec{S}\}$=$[s_1, s_2, ..., s_{\frac{N_{total}}{3}+1}, ..., s_{\frac{2N_{total}}{3}+1}, ..., s_{N_{total}}]$ is the ranking of the subjective scores. N_{total} is the number of the distortion images. For LIVE, LIVEMD, CSIQ, the subjective scores are DMOS, hence these scores are ranked in the ascending order. For LIVEC, TID2013, the subjective scores are MOS, these scores are ranked in the descending order. The quality levels are defined as:

$$L_1 = [s_1, s_{\frac{N_{total}}{3}}] \tag{5}$$

$$L_2 = [s_{\frac{N_{total}}{3}+1}, s_{\frac{2N_{total}}{3}}] \tag{6}$$

$$L_3 = [s_{\frac{2N_{total}}{3}+1}, s_{N_{total}}] \tag{7}$$

where L_1,L_2 and L_3 are the range of the high, medium, low quality level scores, respectively. The distortion images with each quality level are found according to L_1,L_2, L_3.

For different IQA databases, the range of quality scores varies. Therefore, we classify the distortion images into three quality levels, according to the specific MOS/DMOS scale in each IQA database. This can effectively avoid the overlap of quality level semantics caused by uneven score distribution in different IQA databases.

The advantages of this quality level strategy are summarized as follows. First, the quality level strategy is to associate the T_s, T_a and T_t for TTL. This is because it makes T_a a classification task with the quality level, which not only associates T_s based on the classification task with the scene contents, but also associates T_t based on the regression task with the quality semantics. Therefore, the construction of quality level label aims to smooth the transformation from the classification task to the regression task, which progressively enhances the correlation of multiple tasks for TTL.

Moreover, the quality level strategy makes that each distortion image within a quality level also gives the semantic characteristic of human perception. This is because humans prefer to use the natural language to evaluate image quality [47], such as high, medium, low. The natural language metric is a range measure, which reduces the error of absolute scores for different subjects.

In addition, this strategy is beneficial for training the DT-GAN. This is because this strategy reduces the significant difference of the number of images between the high, medium and low levels, which enhances the ability of the D-Dnet to distinguish between the fake $I_d^{'}$ and the real I_d and then enhances the accuracy of the generated quality level label.

C. The loss function of the DT-GAN

Our goal is to train the DT-GAN to generate $I_d^{'}$ with the similar distortion distribution to the IQA database. At the same time, it needs to ensure that the content of $I_d^{'}$ is consistent with the corresponding content of I_hq, which is not affected by the image content of the IQA database. Therefore, the loss function of the DT-GAN is computed by two components: the adversarial loss and the cycle consistency loss.

To make the distortion distribution of $I_d^{'}$ indistinguishable from I_d, we adopt the adversarial loss L_{adv} [45],as:

$$L_{adv}(w;\theta) = L_1(w;\theta) + L_2(w;\theta) \tag{8}$$

$$L_1(w;\theta) = E[logD_w(I_d)] + E[log(1 - D(G_\theta(I_{hq})))] \tag{9}$$

$$L_2(w;\theta) = E[logD_w(I_{hq})] + E[log(1 - D(G_\theta(I_d^{'})))] \tag{10}$$

where $L_1(\cdot)$ is the loss function of the D-Dnet and the D-Gnet; $G(I_{hq})$ aims to generate $I_d^{'}$ in the D-Gnet; The $D(G(I_{hq}))$ tries to discriminate the probability distribution in the D-Dnet between $I_d^{'}$ and I_d. Similarly, $L_2(\cdot)$ is the loss function of the H-Dnet and H-Gnet; $G(I_d^{'})$ aims to generate $I_{hq}^{'}$; The $D(G(I_d^{'}))$ tries to discriminate the probability distribution in the H-Dnet between $I_{hq}^{'}$ and I_{hq}. Finally, the adversarial learning is to minimize L_{adv}, which makes the generated $I_d^{'}$ and $I_{hq}^{'}$ from D-Gnet and H-Gnet to fool D-Dnet and H-Dnet and realize the translation of the distortion distribution.

Although the adversarial loss can translate the distortion distribution into $I_d^{'}$ and preserve its content unchanged, the similarity of statistical characteristics relevant to image content cannot guarantee that $I_{hq}^{'}$ and I_{hq} come from the same image content. Therefore, we propose a cycle consistency loss L_{cyc} to optimize the consistency of image content.

$$L_{cyc}(\theta) = L_p(\theta) + L_s(\theta) \tag{11}$$

$$L_p(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left\|G_\theta(I_d^{'}) - I_{hq}\right\|^2 \tag{12}$$

$$L_s(\theta) = \frac{1}{M}\sum_{i=1}^{M}\left\|\phi(G_\theta(I_d^{'})) - \phi(I_{hq})\right\|^2 \tag{13}$$

where $L_p(\cdot)$ is the pixel-wise loss between $I_{hq}^{'}$ and I_{hq} to represent the holistic content consistency roughly; N is the image dimensions; $L_s(\cdot)$ is the high-level semantic loss to refine local content similarity between $I_{hq}^{'}$ and I_{hq}; $\phi(\cdot)$ represents the extracted high-level semantic features from the last fully connected layer of the VGG for the recognition task; M is the dimensions of the high-level semantic features.

Finally, the objective loss function L is presented to optimize the DT-GAN:

$$L(w;\theta) = L_{adv}(\theta) + \lambda L_{cyc}(\theta) \tag{14}$$

where λ controls the relative importance of the two loss components in the DT-GAN.

D. The training strategy of DT-GAN

The inputs of DT-GAN are the pristine high quality images and the specific quality level IQA images. These pristine high quality images are from the large-scale Waterloo Exploration Database [46], which contains a total of 4744 high quality images diverse content, as shown in the brown box in Fig. 3. The specific quality level IQA images are from the IQA database, as shown in the purple box in Fig. 3. Since the Waterloo Exploration Database includes various scene contents, it can associate the properties of various scene contents among D_s, D_a and D_t.

We train DT-GAN three times. Every time, the inputs are the same 4744 pristine high quality images and the IQA images with a specific quality level label (High, Medium or Low). Considering the insufficient IQA image data, sample enhancement strategies have been applied for training DT-GAN. First, we adopt the distorted samples and their flipped images to enhance the training samples in the IQA database. Second, the training of the discriminator (D-Dnet and H-Dnet) in the DT-GAN is based on PatchGAN method [48]. This method does not send the whole image into the discriminator, instead it divides an image into several patches. The goal is to characterize structures at the image patch scale and classify each $N \times N$ (N=70) patch in an image as real or fake. Therefore, after training the DT-GAN three times, the $I_d^{'}$ with three different quality level labels are generated to construct

TABLE III
THE DIVISION OF QUALITY LEVEL IN EACH IQA DATABASE

Database	Subjective score	Distortion Type	High quality		Medium quality		Low quality	
			Range	Numbers	Range	Numbers	Range	Numbers
LIVEC	MOS	N/A	[67.6 100]	387	[49.4 67.6]	386	[0 49.4]	389
LIVEMD	DMOS	GB+WN	[0 47.5]	75	[47.5 58]	74	[58 100]	76
		GB+JPEG	[0 42]	75	[42 59]	73	[59 100]	77
		JPEG	[0 39]	58	[39 58]	59	[58 100]	58
LIVE	DMOS	JP2K	[0 36]	56	[36 54.5]	56	[54.5 100]	57
		GB	[0 36]	49	[36 50.5]	49	[50.5 100]	47
		WN	[0 35]	49	[35 51]	49	[51 100]	47
		FF	[0 34]	49	[34 51.5]	48	[51.5 100]	48

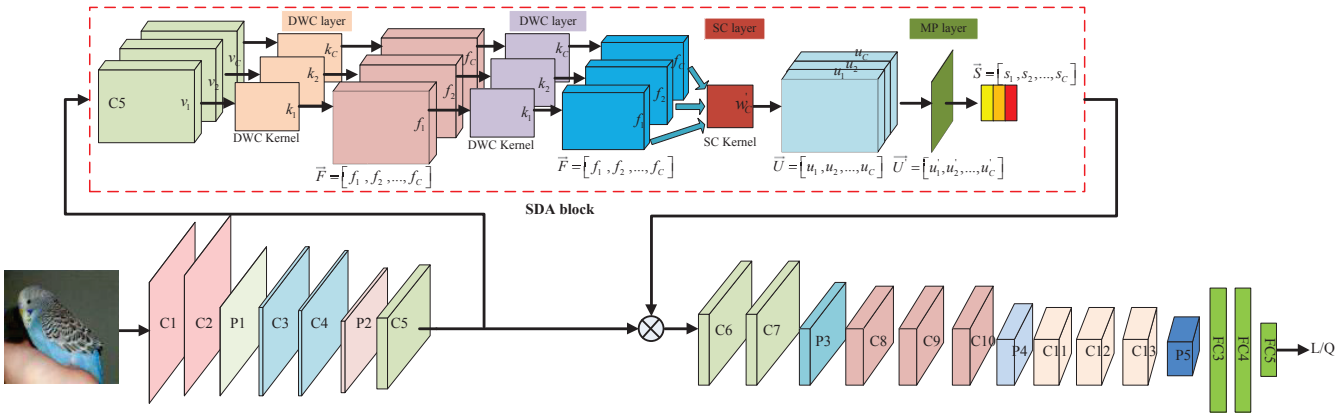

Fig. 6. The SFTnet architecture.

the D_a and the label of T_a. The total number of hallucinated images for each distortion type is 4744×3.

All the training samples are the size of 256×256 pixels that randomly cropped from the high-quality and the distorted images. Generator and the discriminator are trained alternately. First, using $I_d^{'}$ of the D-Gnet, the D-Dnet is optimized to maximize $\underset{D-Dnet}{max}\ L_1$ so that it can correctly discriminate the real I_d and the fake $I_d^{'}$. Then, according to the loss error of the D-Dnet, the D-Gnet can minimize $\underset{D-Gnet}{min}\ L_1$ to fool D-Dnet. This adversarial training ensures that the distortion distribution and the label of the generated $I_d^{'}$ is similar to that of I_d. Then, $I_d^{'}$ are fed into the H-Gnet to generate $I_{hq}^{'}$ so that the H-Dnet can be trained to maximize $\underset{H-Dnet}{max}\ L_2$. It can discriminate the fake $I_{hq}^{'}$ and the real I_{hq}. By adjusting the parameters of H-Gnet, the H-Gnet is also optimized to minimize $\underset{H-Gnet}{min}\ L_2$ and $\underset{H-Gnet}{min}\ L_{cyc}$. The goal is to trick the H-Dnet to judge $I_{hq}^{'}$ as the real I_{hq}. In this way, the image content of $I_d^{'}$ can be preserved from I_{hq}. Finally, the joint optimization of four loss functions in DT-GAN makes the distortion distribution and the label of $I_d^{'}$ be translated from the IQA images without changing their image contents.

V. The SFTnet for TTL-IQA

For TTL, an appropriate TTL network can optimize the shared features among multiple tasks to improve the prediction performance. Although the VGG is used to act as the transfer network to complete the IQA task, not all of the shared features of the salient objects and the highly discriminative and class-specific abilities are useful to the IQA task. Therefore, an appropriate TTL network is important to the TTL-IQA method.

A. The SFTnet architecture

In order to construct an appropriate TTL network, we propose a new SFTnet, which aims to enhance the useful shared features of salient objects and image distortion and suppress the useless shared features of the highly discriminative and class-specific abilities to achieve the IQA task.

Fig. 6 shows the SFTnet architecture. It contains two modules, including the VGG and a new attention unit of the semantic discrimination adaptation (SDA) block. The architecture of SDA block consists of the two depth-wise convolution (DWC) [49], the SC and the maxpooling (MP) layers.

1) The DWC layer of single channel correlation enhancement: In order to accurately identify image semantic properties, we design the two DWC layers, which aim to perform a spatial convolution operation of each channel feature map independently.

For each single channel, the spatial convolution operation is defined as:

$$f_c = v_c \otimes k_c \tag{15}$$

where f_c is the c-th feature map output, which enhances the correlation of spatial locations to highlight the semantic properties. $\otimes$ denotes the element-wise multiplication. v_c

denotes the c-th feature map input of the previous convolution layer. k_c denotes the parameters of the c-th filter in the DWC layer. f_c, v_c and k_c are the 2D size. Especially, $f_c \in \vec{F} = [f_1, ..., f_c, ..., f_C]$. $\vec{F}$ collects the semantic properties of each channel feature map after the DWC operation.

2) The SC layer of the multiple channels correlation enhancement: In order to tackle the issue of discriminating the importance of feature maps over the multiple channels, we design the SC with 1×1 windows, projecting the channels output of the DWC layer into a new channel space with multiple channel convolution operation. The multiple channel output is defined:

$$u_c = \sum_{c=1}^{C} f_c * w_c \tag{16}$$

where u_c is the output of the c-th channel, which aims to establish the interdependency among different channels to highlight local information of an image. $*$ denotes convolution operation. The larger u_c is, the greater the image properties of this channel become. $u_c \in \vec{U} = [u_1, ..., u_c, ..., u_C]$. $\vec{U}$ collects the multiple channel outputs after the SC operation. w_c is the parameters of the c-filter in the SC layer.

3) The MP layer of explicitly establishing the dependency of multiple channels: In order to discriminate the influence of different channels for the IQA task, we explicitly establish the dependency of multiple channels by using the MP layer:

$$u_c^{'} = MaxPool(u_c) \tag{17}$$

where $u_c^{'}$ is the output of the c-th channel by using MP layer. $u_c^{'} \in \vec{U'} = [u_1^{'}, ..., u_c^{'}, ..., u_C^{'}]$. $\vec{U'}$ collects the outputs of the MP layer, which can be interpreted as a collection of the local image information whose statistics are expressive for the whole image.

Then, the sigmoid activation is used to obtain channel weights:

$$s_c = \sigma(\delta(w_c^{'} u_c^{'})) \tag{18}$$

where s_c denotes the weight of the c-th channel, which is to discriminate the importance of local image information from the global image information. δ is the ReLU activation function; σ is the sigmoid function. It means that these weights can selectively emphasize salient object features relevant to the relationship between the recognition and the IQA tasks, as well as the image distortion features relevant to the IQA task. Also, they suppress the features of highly discriminative and class-specific abilities that are not important to the IQA task. $s_c \in \vec{S} = [s_1, ..., s_c, ..., s_C]$. $\vec{S}$ collects the weights of each channel.

Finally, the final outputs of feature maps $\vec{X}$ are obtained by rescaling $\vec{V}$ with the weight activation $\vec{S}$.

$$\vec{X} = \vec{V} \otimes \vec{S} \tag{19}$$

B. The training strategy of the SFTnet

First, before training the SFTnet, the parameters of VGG is shared to the SFTnet. The parameters of the SDA block in the SFTnet is initialized randomly.

Second, the SFTnet is trained from the recognition task to the auxiliary quality level task by using the hallucinated distortion images and the quality level lables. The input images of the SFTnet are randomly cropped to the size of 224×224 pixels. The sofmax cross-entropy loss is used to classify the three quality levels. In this way, the SFTnet is trained to complete the quality level task for TTL.

Third, the SFTnet is trained from the auxiliary quality level task to the IQA quality score task by using the IQA images with the subjective score labels. The last fully-connected layer of the SFTnet is changed to a one-dimensional output. The input IQA images are resized to 224×224 pixels. The Euclidean distance loss between the prediction score and the groundtruth subjective score is used to fine-tune the SFTnet by using the stochastic gradient decent strategy. In this way, the SFTnet is trained to achieve the IQA quality score task for TTL.

VI. Experiments

A. Experimental setups

1) IQA databases: The five public IQA databases are used to evaluate the proposed TTL-IQA method, including LIVE [38], TID2013 [39], CSIQ [50], LIVE multiply distorted (MD) [51] and LIVE In the Wild Image Quality Challenge Database (LIVEC) [37]. Especially, the LIVEC is the authentic IQA database and the LIVE MD is the mixed IQA database. The rest are the synthetic IQA databases. The characteristics of these five IQA databases are summarized in Table IV. Note that Ref means the number of reference images. Dist means the number of distorted images. DT means the number of distortion types. SST and SR denote subjective score's type and range, respectively.

TABLE IV
THE BENCHMARK DATABASES FOR NR-IQA METHODS

Database	Ref.	Dist.	DT	SST	SR
LIVE	29	779	5	DMOS	[0,100]
TID2013	25	3000	24	MOS	[0,9]
CSIQ	30	866	6	DMOS	[0,1]
LIVE MD	15	450	2	DMOS	[0,100]
LIVEC	N/A	1162	Numerous	MOS	[0,100]

2) Evaluation criteria: To evaluate the performance of the TTL-IQA method, we use two standard measures, including Spearman Rank-Order Correlation Coefficient (SROCC) and Pearson Linear Correlation Coefficient (PLCC). The PLCC measures the prediction accuracy and the SROCC measures the prediction monotonicity. For both correlation metrics a value close to 1 indicates high performance of a specific quality measure.

3) Training settings: In the process of training the DT-GAN, we randomly divide the high-quality images into two sets: 80% for training and 20% for testing in the Waterloo Exploration database. The high-quality images in the training sets are used as the inputs of the HD-Gnet. For the IQA database, the distorted images are also randomly divided into two sets in each quality level, 80% of distorted images with

the specific quality level labels are used as the training set for the HD-Dnet and the remaining 20% of distorted images are used for testing. Especially, there is no overlap in the image contents between these training and test sets for the synthetic databases. In the LIVEC database, since all the images are different in content, the training and testing sets are randomly selected. We use the Pytorch framework to train the DT-GAN. We use the Adam solver [52] with a batch size of 1. The learning rate is set as 0.0002. We keep the same learning rate for the first 100 epochs and linearly decay the rate to zero over the next 100 epochs.

Next, we use the Caffe framework to train the SFTnet. The min-batch is set to 30. The momentum and weight decay is set to 0.9 and 0.0005. The learning rate is set to 1e-6. Training rates are decreased by a factor of 0.1 every 10K iterations for a total of 50K iterations. The dropout regularization ration is set to 0.5. Finally, the above the training set of the IQA database is used to fine-tuning the SFTnet.

B. Performance on individual databases

We compare our proposed TTL-IQA method with the state-of-art FR-IQA and NR-IQA methods. The FR-IQA methods contain PSNR, SSIM [2] and FSIMc [53]. The NR-IQA methods contain the classic NR-IQA methods (BLIINDSII [54], BRISQUE [6], BWS [8], CORNIA [55], GMLOG [56] and IL-NIQE [57]), and the deep learning methods (CNN [28], RankIQA [31], BIECON [58], DIQaM [19], DIQA [30], DB-CNN [32], HIQA [33]). Especially, we also compare the TTL-IQA method with the methods that first pre-train a well-known DNN model, such as AlexNet[13], ResNet50 [42] and VGG-16Net [59] for the recognition task and then fine-tune this model for the IQA task.

Since the training and testing sets are randomly selected in our TTL-IQA method, the random process is repeated ten times to eliminate the performance bias. The average results of the obtained SROCC and PLCC values are reported as the final performance. Table V shows SROCC and PLCC on the five public databases. The italics indicate DNN-based methods. The best three results among the NR-IQA methods are shown in bold. The weighted average (WA) of SROCC and PLCC over the five databases is shown in the last column. The weight of each database is proportional to the number of distorted images in the database.

Compared with the performance of the classical NR-IQA methods, the proposed TTL-IQA method is superior over all classical NR-IQA methods. This is because the TTL-IQA method is the deep learning method that can automatically extract deep features relevant to image quality instead of the handcrafted low-level features. Furthermore, compared with the deep learning methods, the proposed TTL-IQA method has the best prediction performance on the authentic distortion database (LIVEC) and the mixed distortion database (LIVE MD). The performance of this method is better than most of other methods on the LIVE, TID2013 and CSIQ.

Compared with the image patches-based methods (CNN, BIECON, DIQaM, DIQA), our TTL-IQA method still outperforms the most of these methods, because the quality labels of image patches used in those methods are inaccurate. In addition, the correlation between image patches was not explicitly considered. Especially, for the LIVEC and LIVE MD databases, the authentic distortions are much more heterogeneous than the synthetic distortions so that it makes less sense to use the label of image patch to describe the image quality. We use the images instead of image patches.

Compared with the external enhancement methods (RANK, DB-CNN, HIQA), the prediction performance of our methods is significantly improved in the LIVE MD and LIVEC. This is because for the LIVEC, the prior distortion information cannot be acquired in advance. For LIVE MD, the influence of mixed distortion types is also complex, which cannot be simply act as the sum of individual distortions. Therefore, these images cannot be artificially simulated, which leads to the inaccurate performance. Furthermore, compared with the methods related to the pre-trained DNN model, our method is better than them in most cases. This is because TTL can learn more shared features than directly transfer learning.

In addition, compared with the FR methods, the TTL-IQA method even outperforms the FR methods in the LIVE and LIVE MD databases. Therefore, the performance of our TTL-IQA method is overall promising.

C. Performance on individual distortion types

To take a closer look at the behaviors of the TTL-IQA on individual distortion types along with several competing NR-IQA methods, the models are trained with all the distortion types from the training set (80%) and tested on individual distortion types from the test set (20%). In Table VI, we compare the performance of the TTL-IQA method and other IQA methods.

As shown in Table VI, even when each distortion type is tested separately, the TTL-IQA method is better than classical and other deep learning NR-IQA methods for most distortion types. We observe the performance of our TTL-IQA method is the best in the LIVE MD database. This is because our proposed method can automatically generate the distortion images in the auxiliary domain, which is similar to the mixed distortion types in the LIVE MD. For the LIVE, CSIQ and TID 2013, our method shows higher prediction accuracy than any other methods on the common distortions (GB, WN, JPEG, JP2K and FF). Especially, for the LIVE and the TID 2013, our method is significantly superior to the other methods for the common FF distortion. Because of the uncertainty of the local distortion, it is difficult to generate the images with this distortion type in [31]. Compared with the FR methods, the TTL-IQA outperforms the FR methods in the LIVE and the LIVE MD. In addition, for the TID 2013, the performance of TTL-IQA on the JGTE, CHA, CTC and Block are even better than the FR methods.

To facilitate a comparison of the performance between the TTL-IQA method and the other NR-IQA methods, there are 37 distortion types in the four databases, including 5 distortion types in LIVE, 24 distortion types in TID2013, 6 distortion types in CSIQ, 2 distortion types in LIVE MD. In the Table VII, we calculate the number of times that the best

TABLE V
THE SROCC AND PLCC COMPARISON ON THE FIVE DATABASES

Types	Algorithms	LIVE		TID2013		CSIQ		LIVE MD		LIVEC		WA	
		SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
FR	PSNR	0.876	0.872	0.636	0.706	0.806	0.800	0.725	0.815	N/A	N/A	N/A	N/A
	SSIM[2]	0.913	0.945	0.775	0.691	0.834	0.861	0.845	0.882	N/A	N/A	N/A	N/A
	FSIMc[53]	0.963	0.960	0.802	0.877	0.913	0.919	0.863	0.818	N/A	N/A	N/A	N/A
NR	BLIINDSSII[54]	0.912	0.916	0.536	0.628	0.780	0.832	0.887	0.902	0.463	0.507	0.628	0.689
	BRISQUE[6]	0.939	0.942	0.572	0.651	0.775	0.817	0.897	0.921	0.607	0.645	0.676	0.729
	CORNIA[55]	0.942	0.943	0.549	0.613	0.714	0.781	0.900	0.915	0.618	0.662	0.659	0.708
	GMLOG[56]	0.950	0.954	0.675	0.683	0.803	0.812	0.824	0.863	0.543	0.571	0.713	0.727
	IL-NIQE[57]	0.902	0.908	0.521	0.648	0.821	0.865	0.902	0.914	0.594	0.589	0.651	0.719
	BWS[8]	0.934	0.943	0.597	0.622	0.786	0.820	0.901	0.922	0.482	0.526	0.666	0.693
	AlexNet[13]	0.942	0.933	0.615	0.668	0.647	0.681	0.881	0.899	0.765	0.788	0.707	0.742
	VGG-16[59]	0.952	0.949	0.612	0.671	0.762	0.814	0.884	0.900	0.753	0.794	0.721	0.765
	ResNet50[42]	0.950	0.954	0.712	0.756	0.876	**0.905**	0.909	0.920	**0.809**	**0.826**	0.797	0.826
	CNN[28]	0.956	0.953	0.558	0.653	0.683	0.754	**0.933**	0.927	0.516	0.536	0.644	0.702
	RANK[31]	**0.981**	**0.982**	0.780	0.793	0.861	0.893	0.908	0.929	0.641	0.675	0.800	0.818
	BIECON[58]	0.961	0.960	0.717	0.762	0.815	0.823	0.909	**0.933**	0.663	0.705	0.765	0.797
	DIQaM[19]	0.960	**0.972**	0.835	0.855	0.869	0.894	0.906	0.931	0.606	0.601	0.800	0.818
	DIQA[30]	**0.970**	**0.972**	**0.843**	**0.868**	0.844	0.880	0.920	**0.933**	0.703	0.704	**0.839**	**0.857**
	DB-CNN[32]	0.968	0.971	0.816	0.865	**0.946**	**0.959**	**0.927**	**0.934**	**0.851**	**0.869**	**0.868**	**0.897**
	HIQA[33]	**0.982**	**0.982**	**0.879**	**0.880**	**0.884**	0.901	–	–	–	–	–	–
	TTL-IQA	**0.979**	**0.984**	**0.844**	**0.869**	**0.895**	**0.907**	**0.952**	**0.960**	**0.884**	**0.890**	**0.884**	**0.899**

Red: the highest performance. Blue: the second best performance. Green: the third best performance.

TABLE VI
THE SROCC COMPARISON ON INDIVIDUAL DISTORTION TYPES

D	T	PSNR	SSIM	FSIMc	BLIINDSII	CORNIA	GMLOG	*RANK*	*BIECON*	*DIQaM*	*TTL-IQA*
LIVE	JP2K	0.895	0.961	0.972	0.930	0.921	0.926	**0.970**	**0.952**	0.944	**0.979**
	JPEG	0.881	0.972	0.979	0.950	0.938	**0.963**	**0.978**	**0.974**	0.928	**0.978**
	WN	0.985	0.969	0.971	0.947	0.957	**0.983**	**0.991**	**0.980**	0.971	**0.983**
	GB	0.782	0.952	0.968	0.915	**0.957**	0.929	**0.988**	0.956	0.936	**0.978**
	FF	0.891	0.956	0.950	0.874	0.906	0.899	**0.954**	**0.923**	0.899	**0.979**
TID2013	AGN	0.934	0.867	0.910	0.714	0.550	**0.748**	0.667	**0.769**	0.512	**0.785**
	ANC	0.867	0.773	0.854	**0.728**	0.209	0.591	0.620	**0.708**	0.313	**0.722**
	SCN	0.916	0.852	0.890	**0.825**	0.717	0.769	0.821	**0.859**	0.744	**0.899**
	MN	0.836	0.777	0.801	0.358	0.360	0.491	0.365	**0.607**	**0.513**	**0.516**
	HFN	0.913	0.863	0.904	**0.852**	0.797	**0.875**	0.760	0.811	0.712	**0.830**
	IN	0.900	0.750	0.825	0.664	0.585	0.693	0.736	**0.753**	**0.760**	**0.780**
	QN	0.875	0.866	0.880	0.780	0.727	**0.833**	0.783	**0.806**	0.783	**0.829**
	GB	0.910	0.967	0.955	0.852	0.840	**0.878**	0.809	**0.882**	0.789	**0.910**
	DEN	0.953	0.925	0.933	0.754	0.721	0.721	**0.767**	**0.780**	0.604	**0.837**
	JPEG	0.922	0.920	0.934	0.808	0.806	0.823	**0.866**	**0.881**	0.762	**0.901**
	JP2K	0.886	0.947	0.959	0.862	0.800	0.872	0.878	**0.902**	**0.899**	**0.921**
	JGTE	0.806	0.845	0.861	0.251	0.595	0.400	0.704	**0.769**	**0.766**	**0.898**
	J2TE	0.891	0.883	0.912	0.755	0.654	0.731	**0.810**	**0.800**	0.717	**0.878**
	NEPN	0.679	0.782	0.794	0.081	0.157	0.190	**0.512**	**0.524**	0.304	**0.496**
	Block	0.330	0.572	0.553	0.371	0.016	0.318	**0.622**	**0.535**	0.226	**0.609**
	MS	0.757	0.775	0.749	**0.159**	0.177	0.119	**0.268**	0.118	**0.344**	**0.222**
	CTC	0.447	0.378	0.468	-0.082	0.262	0.224	**0.613**	0.437	**0.461**	**0.669**
	CCS	0.634	0.414	0.836	0.109	0.170	-0.121	**0.662**	0.044	**0.299**	**0.694**
	MGN	0.883	0.780	0.857	0.699	0.407	0.701	**0.619**	**0.722**	0.469	**0.796**
	CN	0.841	0.857	0.914	0.222	0.541	0.202	**0.644**	0.533	**0.579**	**0.800**
	LCNI	0.916	0.806	0.949	0.451	0.696	0.664	**0.800**	**0.915**	0.599	**0.930**
	ICQD	0.820	0.854	0.882	**0.815**	0.649	**0.886**	0.779	0.807	0.662	**0.848**
	CHA	0.880	0.878	0.893	0.568	**0.689**	**0.648**	0.629	0.609	0.525	**0.913**
	SSR	0.911	0.946	0.958	0.856	**0.874**	**0.915**	0.859	0.626	0.797	**0.942**
CSIQ	WN	0.963	0.896	0.936	0.702	0.763	0.804	**0.844**	0.804	**0.860**	**0.882**
	JPEG	0.888	0.956	0.966	0.846	0.842	0.864	**0.935**	0.752	**0.907**	**0.917**
	JP2K	0.936	0.961	0.970	0.850	0.869	**0.890**	**0.915**	0.837	0.817	**0.924**
	PGN	0.934	0.892	0.937	0.812	0.567	0.774	**0.888**	**0.847**	0.845	**0.913**
	GB	0.929	0.961	0.973	**0.880**	0.854	0.857	0.840	0.822	**0.859**	**0.894**
	CTD	0.862	0.792	0.944	0.336	0.533	0.562	**0.671**	**0.661**	0.592	**0.821**
LIVEMD	GB+JPEG	0.736	0.898	0.885	0.899	**0.900**	0.824	**0.909**	0.797	0.815	**0.919**
	GB+WN	0.743	0.912	0.899	0.892	**0.899**	0.863	**0.933**	0.869	0.812	**0.971**

Red: the highest performance. Blue: the second best performance. Green: the third best performance.

SROCC value appears in the 37 distortion types. Our proposed TTL-IQA method shows 26 times being the best performing method, followed by GMLOG (3 times), BLIINDSII (1 time), CORNIA (0 time), RANK (5 times), BIECON (2 times), DIQaM (1 times). It means the TTL-IQA method is better than other NR-IQA methods in terms of overall performance. Meanwhile, we also report WA and the weighted standard deviation (WSTD) of the competing methods across all distortion groups in Table VII. The TTL-IQA has the highest average and lowest STD across all distortion groups. Hence, the TTL-IQA method achieves a consistently good performance on all available distortion types.

TABLE VII
THE SROCC OF WEIGHTED MEAN AND STD

Algorithms	Weighted Average	Weighted STD
BLINNDSII	0.661	0.285
CORNIA	0.651	0.260
GMLOG	0.686	0.270
RANK	0.763	0.166
BIECON	0.753	0.210
DIQaM	0.703	0.208
TTL-IQA	**0.845**	**0.160**

D. Performance on cross-database test

In the previous experiments, the training and the testing samples are selected from the same database. It is expected that an IQA model that has learned on one image quality database should be able to accurately assess image quality in other IQA databases. Therefore, to demonstrate the generalization ability of the TTL-IQA method, a cross database validation is conducted.

For strict cross-database experiment, when a IQA database is tested, the distortion information of IQA database for test cannot be learned in advance. For our TTL-IQA method, this rule is strictly followed. That is to say, our pre-training method is to train hallucinated images generated by each IQA database, respectively. Therefore, there is a separate pre-training model for each IQA database. When the cross-database experiment measure the IQA database for test, make sure that DNN does not obtain distortion information of IQA database for training. In addition, the compared IQA methods are also follow the strict cross-database experiment.

The SROCC results are shown in Table VIII. We observe that the generalization ability of the proposed TTL-IQA method is better than that of other methods. Also, the generalization ability of deep learning method is better than that of the traditional method (BLINDSII). This is because the deep learning methods can automatically extract the deep features that are highly related to the quality degradation. Compared BIECON and DIQaM methods, the generalization of TTL-IQA method is better than that of BIECON and DIQaM. This is bacause these methods are based on image patch methods. It is difficult to obtain quality label of image patch. Comapred with RANK, our method also has a stronger generalization ability. This is because our method can generate the mixed and the authentic hallucinated distortion images and can adaptively discriminate the useful and useless shared features for TTL.

E. Ablation experiments

In order to evaluate the design rationality of the TTL-IQA method, we conduct several ablation experiments by comparing the proposed network model with several baseline models in various IQA databases. In this ablation experiment, we use the same experimental settings as described in section VI(A).

We design four groups of the comparative experiments, as shown in Table IX. The experiment in the first group is the transfer learning method, which uses the VGG to transfer features from Imagenet source domain to the IQA target domain. The second experiment is to replace the VGG with the designed SFTnet to obtain the prediction performance in the transfer process. The third and the fourth experiments use the DT-GAN to construct the auxiliary domain. However, the difference between the third and fourth experiments is that the former adopts the VGG for TTL, while the latter uses the SFTnet. We observe that our proposed TTL-IQA method achieves the best performance. The experimental result of the first group is the worst, mainly because the Imagenet source domain and the IQA target domain, as well as their tasks are not directly related. The performance of the second group is not better than that of the first group for LIVE and LIVE MD. It may be because the number of training images in these databases is very small, which leads to the overfitting problem by using the SFTnet. However, the performance of the second group is not the best, because it fails to resolve domain irrelevance. For the third group, since the introduction of DT-GAN can enhance training samples, it improves the prediction accuracy. Thus, the performance is also not the best, because it cannot resolve the irrelevance between the recognition and the IQA tasks. Therefore, when the DT-GAN and the SFTnet are used together, the prediction performance is the best among the four groups. This is because our method not only overcomes the shortcoming of the Imagenet source and the IQA target domains by constructing the auxiliary domain and the auxiliary IQA task, but also alleviates irrelevance of the recognition and the IQA tasks by using the SFTnet in TTL.

F. Discussions

1) The rationality of three quality level strategy in the IQA database: In order to show the rationality of three quality level strategy, we have done experiments on the authentic, the mixed, the synthetic IQA databases, as shown in Table X. We observe that the performance of three quality levels is the best. Compared with the two and three quality level strategies, the performance of three quality levels is the best. It may be caused by the increase of the number of images or the increase of the label accuracy. Compared with the three and five quality level strategies, the performance of three quality levels is also the best. This is because although the increasing of image data can enhance the performance for the five quality level strategy, the inaccuracy of the labels can also limits the prediction accuracy.

TABLE VIII
THE SROCC COMPARISON OF THE CROSS-DATABASE TEST

Train	Test	BLIINDSII	BRISQUE	RANK	BIECON	DIQaM	TTL-IQA
LIVE	TID 2013	0.019	0.358	0.518	0.337	0.392	**0.571**
	CSIQ	0.547	0.562	**0.810**	0.710	0.681	0.807
	LIVE MD	0.225	0.301	0.322	0.251	0.275	**0.457**
	LIVEC	0.014	0.337	0.367	0.171	0.111	**0.490**
TID 2013	LIVE	0.130	0.790	0.817	0.628	0.673	**0.832**
	CSIQ	0.105	0.590	0.725	0.663	0.717	**0.749**
	LIVE MD	0.188	0.152	0.273	0.311	0.184	**0.795**
	LIVEC	0.023	0.254	0.276	0.292	0.192	**0.334**
CSIQ	LIVE	0.491	**0.847**	0.710	0.588	0.785	0.840
	TID 2013	0.019	0.454	0.477	0.350	0.464	**0.497**
	LIVE MD	0.300	0.296	**0.396**	0.255	0.275	0.375
	LIVEC	0.052	0.131	0.265	0.238	0.200	**0.298**
LIVE MD	LIVE	0.511	0.681	0.852	0.697	0.612	**0.880**
	TID 2013	0.013	0.255	0.531	0.429	0.368	**0.585**
	CSIQ	0.574	0.501	0.851	0.598	0.661	**0.863**
	LIVEC	0.049	0.062	0.179	0.300	0.241	**0.314**
LIVEC	LIVE	0.010	0.238	–	–	0.315	**0.676**
	TID 2013	0.133	0.280	–	–	0.198	**0.338**
	CSIQ	0.096	0.241	–	–	0.333	**0.606**
	LIVE MD	0.112	0.355	–	–	0.462	**0.780**

TABLE IX
THE SROCC FOR BASELINE MODELS

Groups	DT-GAN	SFTnet	LIVE	CSIQ	LIVEC	LIVEMD
1	×	×	0.952	0.762	0.753	0.884
2	×	√	0.943	0.785	0.774	0.846
3	√	×	0.962	0.812	0.878	0.918
4	√	√	**0.979**	**0.895**	**0.884**	**0.952**

TABLE X
THE SROCC OF DIFFERENT QUALITY LEVELS IN DIFFERENT DATABASES

Quality levels	LIVE	CSIQ	LIVEC	LIVEMD
No-level	0.952	0.762	0.753	0.884
Two-levels	0.963	0.817	0.875	0.926
Three levels	**0.979**	**0.895**	**0.884**	**0.952**
Five levels	0.960	0.811	0.870	0.901

TABLE XI
THE QUALITY SCORES OF HALLUCINATED IMAGES CALCULATED BY OBJECTIVE IQA METRICS, SSIM AND BRISQUE.

Index	SSIM [2]			BRISQUE [6]		
	High	Medium	Low	High	Medium	Low
Fig. 7(a)	0.831	0.803	0.731	0.678	0.463	0.347
Fig. 7(b)	0.879	0.778	0.684	0.694	0.524	0.327

TABLE XII
THE QUALITY SCORE AVERAGED OVER ALL HALLUCINATED IMAGES (PER LEVEL) CALCULATED BY SSIM AND BRISQUE METRICS FOR THE LIVEC DATABASE

Quality levels	High	Medium	Low
Average SSIM score	0.798	0.745	0.699
Average BRISQUE score	0.622	0.498	0.374

Therefore, considering the prediction performance in the synthetic, mixed and authentic IQA databases, we select the three quality level strategy to generate hallucinated distortion images as a compromise.

2) The accuracy of quality level of hallucinated distortion images: We show the hallucinated distortion images of different quality levels, as shown in Fig. 7. The images are derived from the Waterloo Exploration Database whose index is 250 and 415, respectively. We use the LIVEC database to generate the hallucinated distortion images. To help readers easily see the differences in quality, the distortion portion is highlighted according to the method in [61]. The hallucinated distortion image in Fig. 7 is cropped out a portion of an hallucinated distortion image and illustrate that portion closer to the original resolutions. We observe the quality level can be discriminated. In order to explicitly represent quality level discrimination, we also use SSIM [2] and BRISQUE [6] to calculate the difference quality score, as shown in Table XI. We observe these hallucinated distortion images with different quality levels can easily be discriminated by using different quality scores. It should be noted, the purpose of using SSIM and BRISQUE, as objective quality indicators, in this part of this paper is only to provide an supplementary tool (in additional to the visual inspection as mentioned above) to help confirm the distinctive quality of levels. Moreover, we show the objective quality score averaged over all hallucinated images for each level using SSIM and BRISQUE on the LIVEC database, as shown in Table XII. This clearly shows that the hallucinated images reflect three distinctive levels of perceived quality.

3) Comparison of different generation methods: In order to ensure that the DT-GAN can promote the prediction performance for the IQA, we compare the different generation methods, including the DT-GAN and the artificial simulation. Especially, The artificial simulation method is to generate the synthetic distortion images with three quality levels from the RANK method [31]. Since there is lack of the prior

TABLE XIII
THE SROCC AND PLCC OF DIFFERENT GENERATION METHODS

Methods	LIVEC		LIVE		CSIQ		LIVE MD		TID2013	
	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
DT-GAN	0.884	0.890	0.979	0.984	0.847	0.888	0.952	0.960	0.821	0.848
Artificial simulation	0.796	0.802	0.962	0.946	0.810	0.832	0.935	0.945	0.770	0.779

distortion information in the LIVEC database, the prediction performance is obtained by training the model of the TID2013 database. In order to ensure fairness, after expanding the image data, we use the SFTnet to pre-train these distortion images and then fine-tune the IQA database. The prediction performance is presented in Table XIII. We observe that the DT-GAN is better than the artificial simulation method [31]. This is because DT-GAN can easily simulate various distortion types, including synthetic, mixed and authentic distortions. Also, DT-GAN can learn the distribution of distortion properties of the IQA databases, making sure the simulated distortions have a similar distribution to that of the image distortions in the IQA databases.

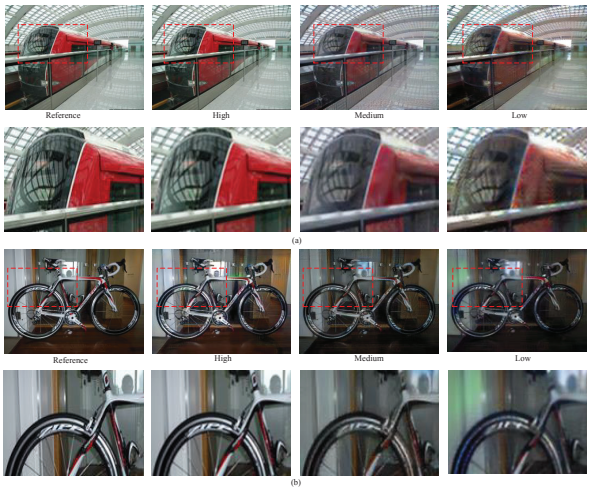

Fig. 7. The different quality levels of a hallucinated distortion image. (a) Results of the 250th image of the Waterloo database. (b) The 415th image of the Waterloo database.

4) Comparison of different authentic IQA databases: In the previous experiments, the performance of our TTL-IQA method is significantly improved in the authentic LIVEC database. It is necessary to verify the prediction performance of our TTL-IQA method in the large scale authentic IQA database (KonIQA-10k [63]). However, it should be noted that directly testing the current model on KonIQA-10k is considered inappropriate. This is mainly because the generated auxiliary domain currently contains a total of 14, 232 hallucinated images, which is disproportional to the KonIQA-10k database (10, 073 images) being used as the test set. More specifically, if we were to conduct the experiment with above setting, the number of images in the pre-training and fine-tuning stages would be approximately the same, which would violate the conventional train-test split requirements of transfer learning [16]. Therefore, we should expect that by increasing the number of hallucinated images in the auxiliary set, the model's performance testing on KonIQA-10k would increase.

To verify this hypothesized trend, we now check how the model performance changes with the increase in the number of hallucinated images in the auxiliary domain, as the results shown in Table XIV. Note TTL-IQA is our current model in the manuscript; and TTL-IQA1 and TTL-IQA2 represent the modified models by adding more hallucinated images for training. In this new experiment, based on the PASCAL VOC 2012 database [64] with 17,125 images of diverse content, we randomly select some images to generate additional hallucinated images. As can be seen from Table XIV that with the increase in the number of hallucinated images, the model's prediction performance on KonIQA-10k indeed increases. Scaling up the process is straightforward and would eventually make the auxiliary set proportional to the test set of KonIQA-10k, but this would have involved massive experimental efforts, which will be explored in the next further work.

TABLE XIV
The prediction performance of TTL-IQA on the KonIQA-10k database. TTL-IQA1 and TTL-IQA2 represent modified models with a increased number of hallucinated images for training.

Method	Number of hallucinated images	SROCC	PLCC
TTL-IQA	14232(orginal)	0.691	0.708
TTL-IQA1	18732	0.713	0.722
TTL-IQA2	29232	0.755	0.770

5) Comparison of FR-based proxy score method: In addition to the above quality level as proxy label, we also compare the performance of our method and various FR-based proxy score methods on the synthetic, mixed and authentic IQA databases, as shown in Table XV. The BIECON method [58] uses an FR metric to generate image patch proxy quality scores. Note to be able to evaluate BIECON on LIVEC (pristine reference images are unavailable so BIECON cannot be directly applied as mentioned above), the model is trained on the TID2013 database, with the purpose of making the domain closer to the target LIVEC database. The SSIM-label and VIF-label methods use SSIM [2] and VIF [62] as FR metrics to calculate image proxy scores in the auxiliary domain. Note we generate hallucinated images using the pristine images contained in the Waterloo Exploration Database [46]. A traditional simulation method [31] is used to simulate four distortion types (i.e., JPEG, JP2K, WN, GB contained in the LIVE database). The SSIM and VIF metrics are used to calculate proxy scores of hallucinated images, respectively. As can be seen from Table XV, the prediction performance of these FR-based methods is lower than our TTL-IQA method. It suggests that the domain problem still exists for the FR-

based methods. This might be due to the fact that the proxy quality scores generated by the FR-based methods cannot sufficiently reflect the distortion properties of the target IQA databases, so the features learned in the pre-training stage are less relevant/accurate for quality prediction.

TABLE XV
Comparison of performance of our method vs FR-based proxy score methods on the synthetic (LIVE), mixed (LIVE MD) and authentic (LIVEC) IQA databases

Method	LIVE		LIVEMD		LIVEC	
	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
SSIM-label	0.930	0.940	0.903	0.918	0.824	0.857
VIF-label	0.945	0.934	0.914	0.925	0.826	0.847
BIECON	0.961	0.960	0.909	0.933	0.663	0.705
TTL-IQA	**0.979**	**0.984**	**0.952**	**0.960**	**0.884**	**0.890**

6) Discrimination of the feature importance in SFTnet: While the SFTnet has been shown to improve the DNN performance, we would also like to directly know how to discriminate the feature importance. Thus, two images with WN (Fig. 8(a)-(b)) are fed into the SFTnet, respectively. Fig. 8(a)-(b) are the global and local distortion images with WN. Then, we extract the visualization of the feature maps and their corresponding weights from the SFTnet. Fig. 8(c)-(d) show the visualization of the 129th and the 210th feature maps of the global distortion image. Fig. 8(e)-(f) show the visualization of the 183th and the 186th feature maps of the salient object distortion image. We observe that whether these images represent global or salient object distortion, the SFTnet can emphasize the salient objects and the distortion features that are important to the IQA task by assigning large weights to the feature maps. The features of the highly discriminative and class-specific abilities that are not important to the IQA task are suppressed by assigning small weights to the feature maps.

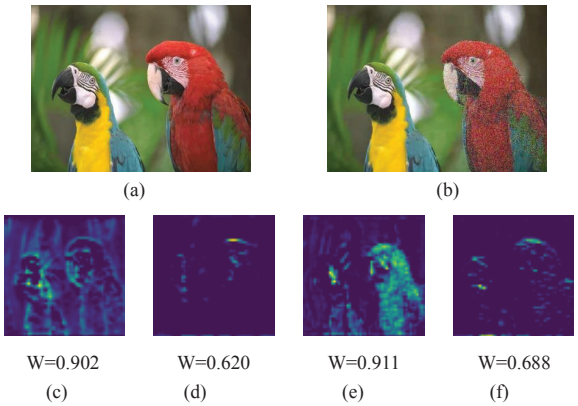

Fig. 8. The visualization of the different feature maps from SFTnet.

VII. Conclusion

In this paper, we propose a new TTL-IQA method for IQA. First, we develop a TTL architecture to enhance the correlation between multiple domains and between multiple tasks. Moreover, we propose a DT-GAN to construct the auxiliary domain and the auxiliary quality level task, which acts as the intermediate bridge for TTL. Finally, a newly proposed SFTnet is used in the TTL network, which optimizes the utilization of the shared features for the IQA task. Experiments demonstrate that the proposed TTL-IQA method is superior over alternative NR-IQA methods in most cases. Especially for the authentic distortion (LIVEC) and the mixed distortion (LIVE MD) images. Moreover, the TTL-IQA method also shows a strong generalization ability.

In addition, some future work is proposed to solve the IQA task by using GAN. We hope this work could help researchers to design effective IQA methods and to foster many potential applications.

First, the perceived quality score label of the hallucinated distortion image generated by GAN could be studied. If the IQA images can be assigned to the perceived quality score during training GAN, it will be more beneficial to optimize the final IQA quality score task.

Second, Multi-distortion types are selected simultaneously to generate the hallucinated distortion images. After training GAN only once, it could automatically complete the translation of distortion distribution under different distortion types. This has a good potential for the enhancement of the IQA database.

Third, it is worth studying the loss function of GAN. The loss function of GAN is considered to complete the translation of distortion distribution. Moreover, the different constraints of the loss function need to be considered to control the performance of the hallucinated distortion image. In addition, the matching degree among the loss functions of the whole GAN could also be studied to benefit the fast convergence and the final generation effect.

References

[1] F. Li, S. Fu, Z. Li and X. Qian, "A cost-constrained video quality satisfaction study on mobile device," *IEEE Trans. Multimedia*, vol. 20, no. 5, pp. 1154–1168, May. 2018.

[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[3] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency-induced index for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4270–4281, Aug. 2014.

[4] S. Wang, K. Gu, X. Zhang, W. Lin, L. Zhang, S. Ma, and W. Gao, "Subjective and objective quality assessment of compressed screen content images," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 6, no. 4, pp. 532–543, Dec. 2016.

[5] Y. Liu, G. Zhai, K. Gu, X. Liu, D. Zhao, and W. Gao, "Reduced-reference image quality assessment in free-energy principle and sparse representation," *IEEE Trans. Multimedia*, vol. 20, no. 2, pp. 379–391, Feb. 2018.

[6] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, 2012.

[7] Q. Wu, H. Li, F. Meng, K. N. Ngan, B. Luo, C. Huang, and B. Zeng, "Blind image quality assessment based on multichannel feature fusion and label transfer," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 3, pp. 425–440, Mar. 2016.

[8] X. Yang, F. Li, W. Zhang, and L. He, "Blind image quality assessment of natural scenes based on entropy differences in the DCT domain," *Entropy*, vol. 20, no. 12, pp. 885–906, 2018.

[9] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu and D. David, "Blind image quality assessment based on high order statistics aggregation," *IEEE Trans. Image Process.*, pp. 4444–4457, 2016.

[10] Q. Jiang, F. Shao, W. Lin, K. Gu, G. Jiang and H. Sun, "Optimizing Multi-Stage Discriminative Dictionaries for Blind Image Quality Assessment," *IEEE Trans. Multimedia*, pp. 2035–2048, 2018.

[11] Q. Jiang, F. Shao, W. Gao, Z. Chen, G. Jiang and Y. Ho, "Unified no-reference quality assessment of singly and multiply distorted stereoscopic images," *IEEE Trans. Image Process*., pp. 1866–1881, 2019.

[12] X, Yang, F. Li and H. Liu, "A study of DNN methods for blind image quality assessment," *IEEE Access*, pp. 123788-123806, 2019.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *in Advances in Neural Information Processing Systems*, pp. 1097–1105, 2012.

[14] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," *in Proc. CVPR*, 2009.

[15] J. Kim, H. Zeng, D. Ghadiyaram, S. Lee, L. Zhang and A. C. Bovik, "Deep convolutional neural models for picture-quality prediction: Challenges and solutions to data-driven image quality assessment," *IEEE Signal processing magazine*, pp. 130–141, 2017.

[16] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowledge and Data Engineering*, pp. 1345–1359, 2010.

[17] H. Yi, Z. Xu, Y. Wen and Z. Fan, "Multi-source domain adaptation for face recognition," *in Proc. ICPR*, pp. 1349-1354, 2018.

[18] Y. Li, L. M. Po, L. Feng, and F. Yuan, "No-reference image quality assessment with deep convolutional neural networks," *in Proc. IEEE Int. Conf. Digital Signal Processing*, pp. 685–689, 2016.

[19] S. Bosse, D. Maniry, K. R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Trans. Image Process*., vol. 27, no. 1, pp. 206–219, Jan. 2018.

[20] B. Gong, K. Grauman, and F. Sha, "Reshaping visual datasets for domain adaptation," *in Proc. 27th Annu. Conf. Neural Inf. Process. Syst*., pp. 1286–1294, 2013.

[21] Y. Chen, W. Li, C. Sakaridis, D. Dai, L.V. Gool, "Domain adaptive faster r-cnn for object detection in the wild," *in Proc. CVPR*, pp. 3339-3348, 2018.

[22] L. Lucie, W. Zhang, H. Liu, "Subjective assessment of image quality induced saliency variation," *in Proc. ICIP*, 2019.

[23] W. Zhang, R. Martin and H. Liu, "A saliency dispersion measure for improving saliency-based image quality metrics," *IEEE Trans. Circuits Syst. Video Technol*., vol. 28, no. 6, pp. 1462–1466, Jul, 2018.

[24] R. Xu, Z. Chen, W. Zuo, J, Yan, L. Lin, "Deep cocktail network: multi-source unsupervised domain adaptation with category shift," *in Proc. CVPR*, pp.3964-3973, 2018.

[25] B. Tan, Y. Song, E. Zhong and Q. Yang, "Transitive transfer learning," *in Proceedings of the 21th ACM SIGKDD International Conference of Knowledge Discovery and Data Mining*, pp. 1155-1164, 2015.

[26] J. Zhang, W. Zhou, X. Chen, W. Yao and L. Gao, "Multi-source selective transfer framework in multi-objective optimization problems," *IEEE Transactions on Evolutionary Computation*, 2019. [Oline]. Available: http://doi.org/10.1109/TEVC.2019.2926107

[27] P. Mohammadi, A. Ebrahimi-Moghadam, and S. Shirani, "Subjective and objective quality assessment of image: A survey," *Majlesi J. Elect. Eng*., vol. 9, no. 1, pp. 55–83, Mar. 2015.

[28] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," *in Proc. CVPR*, pp. 1733–1740, 2014.

[29] B. Bare, K. Li, and B. Yan, "An accurate deep convolutional neural networks model for no-reference image quality assessment," *in Proc. ICME*, pp. 1356–1361, 2017.

[30] J. Kim, A. Nguyen, and S. Lee, "Deep CNN-based blind image quality predictor," *IEEE Trans. Neural Netw. Learn. Syst*., vol. 30, no. 1, pp.11–24, 2019.

[31] X. Liu, J. van de Weijer, and A. D. Bagdanov, "RankIQA: Learning from rankings for no-reference image quality assessment," *in Proc. ICCV*, pp. 1040–1049, 2017.

[32] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, "Blind image quality assessment using a deep bilinear convolutional neural network," *IEEE Trans. Circuits Syst. Video Technol*., vol. 30, no. 1, pp. 36-47, Jan. 2020.

[33] K.-Y. Lin and G. Wang, "Hallucinated-IQA: No-reference image quality assessment via adversarial learning," *in Proc. CVPR*, pp. 732-741, 2018.

[34] S. Chopra, R. Hadsell, and Y. Cun, "Learning a similarity metric discriminatively with application to face verification," *in Proc. CVPR*, pp. 349–356, 2005.

[35] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow and B. Frey, "Adversarial autoencoders," *in Proc. ICLP*, 2016.

[36] P. Isola, J. Y. Zhu, T. Zhou and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *in Proc. CVPR*, pp. 1125–1134, 2017.

[37] D. Ghadiyaram and A. C. Bovik, "Massive online crowd-sourced study of subjective and objective picture quality," *IEEE Trans. Image Process*., vol. 25, no. 1, pp. 372–387, Jan. 2016.

[38] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process*., vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[39] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. J. Kuo, "Image database TID2013: Peculiarities, results and perspectives," *Signal Process., Image Commun*., vol. 30, pp. 57–77, Jan. 2015.

[40] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," *in Proc. ECCV*, 2014.

[41] K. Fu, Q. Zhao, I. Y. Hua, "Refinet: A deep segmentation assisted refinement network for salient object detection," *IEEE Trans. Multimedia*, vol. 19, no. 11, pp. 2505—2521, 2017.

[42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *in Proc. CVPR*, pp. 770–778, 2016.

[43] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," *in Proc. ECCV*, pp. 184–199, 2014.

[44] A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts," *Distill*, 2016. [Online]Available: http://distill.pub/2016/deconv-checkerboard.

[45] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative adversarial networks," *in Proc. CoRR*, 2014.

[46] K. Ma, Z. Duanmu, Q. Wu, Z. Wang, H. Yong, H. Li, and L. Zhang, "Waterloo Exploration Database: New challenges for image quality assessment models," *IEEE Trans. Image Process*., vol. 26, no. 2, pp. 1004–1016, Feb. 2017.

[47] W. Hou, X. Gao, D. Tao, and X. Li, "Blind image quality assessment via deep learning," *IEEE Trans. Neural Netw. Learn. Syst*., vol. 26, no. 6, pp.1275–1286, 2015.

[48] P. Isola, J.Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *in Proc. CVPR*, 2017.

[49] L. Kaiser, A. N. Gomez, and F. Chollet, "Depthwise separable convolutions for neural machine translation," 2017, arXiv:1706.03059. [Online]. Available: https://arxiv.org/abs/1706.03059.

[50] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag*., vol. 19, no. 1, pp. 19–21, 2010.

[51] D. Jayaraman, A. Mittal, A. K. Moorthy, and A. C. Bovik, "Objective quality assessment of multiply distorted images," *in Proc. Asilomar Conf. Signals, Syst. Comput*., pp. 1693–1697, 2012.

[52] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *in Proc. ICLR*, 2015.

[53] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process*., vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[54] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process*., vol. 21, no. 8, pp. 3339–3352, Aug.2012.

[55] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," *in Proc. CVPR*, pp. 1098–1105, 2012.

[56] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features," *IEEE Trans. Image Process*., vol. 23, no. 11, pp. 4850–4862, Nov. 2014.

[57] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process*., vol. 24, no.8, pp. 2579–2591, Aug. 2015.

[58] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE J. Sel. Topics Signal Process*., vol. 11, no. 1, pp. 206–220, Feb. 2017.

[59] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *in Proc. ImageNet Challenge*, pp. 1-14, 2014.

[60] W. L. Hou and X. B. Gao, "Saliency-guided deep framework for image quality assessment," *IEEE Multimedia Mag*., vol. 22, no. 2, pp. 46–55, 2015.

[61] W. Zhang, H. Liu, "Learning picture quality from visual distraction: psychophysical studies and computational models," *Neurocomputing*, pp. 183–191, 2017.

[62] H. Sheikh, and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process*., vol. 15, pp. 430–440, 2006.

[63] H. Lin, V. Hosu, and D. Saupe, "Koniq-10k: Towards an ecologically valid and large-scale iqa database," *CoRR*, 2018.

[64] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," *Int. J. Comput. Vis*., vol. 88, no. 2, pp. 303–338, Jun. 2010.

Xiaohan Yang obtained her M.S. degree in Information Engineering and Automation from Kunming university of science and technology, Kunming, China, in 2016. She is currently pursuing the Ph.D. degree at School of Electronic and Information Engineering, Xi'an Jiaotong University. Her research interests mainly focus on image quality assessment.

Fan Li (M'10) obtained his B.S. and Ph.D. degrees in information engineering from Xi'an Jiaotong University, Xi'an, China, in 2003 and 2010, respectively. From 2017 to 2018, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, University of California, San Diego. He is currently a Professor with the School of Electronic and Information Engineering, Xi'an Jiaotong University. He has published more than 30 technical papers. His research interests include multimedia communication and video quality assessment. He served as the Local Chair for ICST Wicon 2011, and was a member of the Organizing Committee for IET VIE 2008.

Hantao Liu received the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands in 2011. He is currently an Associate Professor with the School of Computer Science and Informatics, Cardiff University, Cardiff, U.K. He is an Associate Editor of the IEEE Transactions on Human-Machine Systems and the IEEE Transactions on Multimedia.